2009

On model validation for meso/submesoscale currents: Metrics and application to ROMS off Central California

Ivanov, L.M.

Ocean Modelling 28 (2009) 209–225

Contents lists available at ScienceDirect

Ocean Modelling

journal homepage: www.elsevier.com/locate/ocemod

On model validation for meso/submesoscale currents: Metrics and application to ROMS off Central California

L.M. Ivanov [a,*], C.A. Collins [b], P. Marchesiello [c], T.M. Margolina [b,d]

[a] Moss-Landing Marine Laboratories, Moss-Landing, CA 95033, USA
[b] Naval Postgraduate School, Monterey, CA 93940, USA
[c] IRD, Centre de Bretagne, Plouzane 29280, France
[d] Marine Hydrophysical Institute, Sevastopol 99011, Ukraine

ARTICLE INFO

Article history:
Received 8 October 2008
Received in revised form 4 February 2009
Accepted 7 February 2009
Available online 20 February 2009

Keywords:
Models
ROMS
Submesoscale features
Mesoscale features
Subsurface drifters
Satellite altimetry

ABSTRACT

ROMS with horizontal grid spacing of 3.5 km for the region off Central California was compared to RAFOS float observations and satellite altimetry on meso/submesoscales. The approach introduced and used two new metrics for model-data comparison, as well as suggested how to calculate these metrics for different spatio-temporal scales. The first metric consisted of the first two moments of exit time and was used to compare ROMS against RAFOS float observations at mid-depths (between 300 m and 350 m). Exit time is the time a float launched at a point takes to leave a domain for the first time. The second metric was spectral entropy and was used to estimate how well ROMS reproduced variability of the sea surface height (SSH) anomaly field extracted from an AVISO data set (1992–2007) for specified temporal and spatial scales. Calculations showed that ROMS reproduced the mid-depth mesoscale/submesoscale currents next to the coast in a very accurate manner (low-order exit time statistics of floats were reproduced by ROMS with an accuracy better than 95%); but ROMS overestimated the speed of westward drift of floats by as much as 20–30% at distances greater than 350 km from the coastline. ROMS predicted the variability of the mesoscale (100–400 km) SSH anomaly field for temporal scales of 1–12 months with a reasonable accuracy. A wavelet transform modulus maxima technique applied to the spectral entropy of SSH anomaly also demonstrated good agreement between ROMS and satellite altimetry for mesoscales characterized by singular exponents and multi-fractal spectra for 1–12 month time scales.

1. Introduction

Currents along ocean eastern boundaries have a broad range of temporal and spatial scales and include waves and wave-eddy structures, meso- and submesoscale eddies, tides, and inertial motions. Progress towards better understanding of this variability is closely linked to the development of (a) high-resolution models with good physical parameterizations which are able to predict coastal flows on time scales ranging from days to months, and (b) adequate signal processing tools for multi-scale analysis of predicted flows characterized by a hierarchy of spatial and temporal scales. The latter is very important for the analysis of high-resolution model simulations because understanding of three-dimensional (3D) mesoscale and submesoscale features (eddies, fronts, filaments, etc.) and their variability will most likely come from three-dimensional numerical simulations at progressively finer resolution (Capet et al., 2008, among others). Clearly, these high-resolution models need to be validated using all available observations of mesoscale/submesoscale features collected in a region of interest.

A high-resolution model provides forecasts of flows that are not fully replicated by *in-situ* data within a range of variables, parameters or scales. How can results of these models be compared to ocean observations? Opposite views are expressed in modern scientific literature. Oreskes et al. (1994) noted the impossibility of validation of complex models because "natural systems are never closed and model results are always non-unique." This reinforces the view that direct comparison of ocean models and observations is difficult since neither the dynamics of numerical models nor the model input (bathymetry, external forcing and subscale parameterizations) are identical to nature. This view is popular among oceanographers and geophysicists (for example, see Wunsch and Heimbach, 2007, among others), and it allows only for model verification (model verification is done to ensure that the model is programmed correctly, algorithms have been implemented properly, and the model does not contain errors, oversights, or bugs). But without correct physics and clearly formulated requirements for what the model should reproduce, "impossibility of validation" (Oreskes et al., 1994) seems strained. There are many examples demonstrating that forecasting

* Corresponding author. Tel.: +1 831 656 3257.
E-mail address: lmivanov@nps.edu (L.M. Ivanov).

1463-5003/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ocemod.2009.02.003

may simply depend on finding the right level for describing the system. It is essential to determine if the model-data differences come from deficiency in modeling ocean physics, from some unessential imperfection, or from unrepresentative data.

Another point of view (for example, see Sornette et al., 2007) is that because our goal is to predict coarse-grained properties of a flow, model-data comparison is possible only for coarse-grained scales which should be predicted from finer scale modeling. This should be achieved through the use of appropriate metrics and for specified spatio-temporal scales.

Three possible scenarios of model-data comparison can be introduced (Chu et al., 2004a). First, the model reproduces the pattern of the real circulation attractor (the attractor is a robust dynamical regime of a flow) including its small-scale details, and may also predict flow behavior for long time intervals (global predictability). The model is comparable to observations at any spatio-temporal scales and for long time periods. This situation is probably impossible to achieve in oceanographic practice, and the "impossibility" statement from Oreskes et al. (1994) is applicable here.

Second, the circulation attractor is correctly reproduced but not all small-scale details of the circulation topology and scales are resolved. Here the predictability of exact flow dynamics exists only for short and intermediate time intervals (local predictability). However, the model is able to predict coarse-grained properties of the flow in the statistical sense. This case is most typical in oceanographic practice and here model-data comparison is possible for specified spatio-temporal scales and specified properties using a number of appropriate metrics.

Third, the predicted attractor pattern differs from the observed one even in large-scale details and the model approximates the flow only within very short time intervals (partial predictability). Data-model comparison is also possible for this case but only for short time intervals.

This paper develops an approach for comparison of high-resolution model output against different types of data for the last two scenarios. New metrics are introduced for comparison on meso/submesoscales, and a technique is suggested to calculate these metrics for specified spatio-temporal scales. This approach is then used to validate the meso/submesoscale variability produced by ROMS off Central California (Marchesiello et al., 2003). Earlier, Centurioni et al. (2008) analyzed the same ROMS configuration but with larger horizontal grid spacing of 5 km and in a larger domain; the time-averaged circulation and associated eddy energy were used as metrics for comparison and the ability of the model to produce four permanent large-scale flow meanders was determined. The present study analyses the ability of ROMS to reproduce smaller temporal and spatial scales.

The rest of the paper is organized as follows. Section 2 introduces metrics for the ROMS-data comparison and explains their benefits. A double (both spatial and temporal) spectral analysis for sea surface height (SSH) anomaly field reproduced by ROMS and extracted from satellite altimeter observations is discussed in Section 3. Section 4 describes observations and ROMS output fields used for the ROMS-data comparison. Sections 5 and 6 describe results of ROMS–RAFOS float data and ROMS-satellite altimetry data comparisons. Conclusions are given in Section 7. Boundary conditions for calculation of basis functions (modes) used in the double spectral analysis are discussed in Appendix A. Appendix B explains a multi-fractal formalism for the ROMS-satellite altimetry data comparison.

2. Metrics for model-data comparison

The quality of model simulations relative to observations is usually assessed by means of basic statistics such as bias, root mean square error (RMSE), correlation coefficients and some skill scores [see Joliffe and Stephenson (2003) for a general review]. The present study also uses RMSE to quantify model accuracy for reproduction of ocean currents but only within specified spatio-temporal scales. This differs from the traditional ocean model-data comparison and allows for better understanding the contribution of different scales to model predictability.

There are recent attempts to apply more advanced techniques for model-data comparison, e.g. a Bayesian approach (see Furrer et al., 2007, as an example). However, the Bayesian analysis does not take into account non-linear characteristics of the data which are important for understanding dynamic processes. That is why the Bayesian approach is not used in the present study.

In general RMSE or correlation coefficients are not robust to observational sampling because the length of sampling is typically not long and observational errors may not be small. This means that the model-data comparison is within observational error. RAFOS float data is an example of this situation. To compare ROMS against RAFOS float observations, new metrics are introduced and used here. These metrics are mean exit time $\langle T\rangle$ and its variance $\langle \delta T^2\rangle$. Following Gardiner (2004), the exit time T is the length of time a float launched at point $\mathbf{X}_o$ takes to leave a domain Ω for the first time:

$$T(\mathbf{X}_o) = \inf_t (t > 0 | \mathbf{X}_o,\ \mathbf{X} \notin \Omega \cup \partial\Omega), \tag{1}$$

where $\notin$ denotes that a point is outside of domain Ω or its boundary $\partial\Omega$, $\mathbf{X}$ are float coordinates, $\cup$ shows that domain of interest included both Ω and $\partial\Omega$. Oceanographic applications of the mean exit time are given by Artale et al. (1997), La Casce and Bower (2000), Lacorata et al. (2001), d'Ovidio et al. (2004), among others. Both $\langle T\rangle$ and $\langle \delta T^2\rangle$ were used by Ivanov et al. (1994) and Ivanov and Chu (2007a,b) to quantify Lagrangian and Eulerian predictability. These characteristics are robust to variations in observational sampling or small statistical ensembles (Ivanov and Chu, 2007a,b). This allows calculating the exit time when the number of floats in a region of interest is not large. Another explicit benefit of $\langle T\rangle$ and $\langle \delta T^2\rangle$ is that they can be used as criteria to distinguish diffusive from advective processes (Ivanov et al., 2008).

Historically, the float means and variances are widely used for the analysis of Lagrangian data [see examples in Griffa et al. (2007)]. However exit time seems to be preferable to these statistics when the sample size is small. Note also the traditional means and variances are easily calculated from knowledge of the exit time statistics (Gardiner, 2004). These arguments support use of the exit time moments instead of the float means and variances.

The spectral analysis from Ivanov and Collins (in press) was used to decompose the SSH anomaly field within single-connected (without islands or a disconnected boundary) domains:

$$\zeta(\mathbf{x}, t) = \sum_{k=1}^{K} a_k(t)\psi_k(\mathbf{x}), \tag{2}$$

where $\mathbf{x}$ is the horizontal space coordinate, the basis functions ψ_k (also called M-modes) are eigenfunctions of the plane Laplace operator for mixed boundary conditions (see Appendix A for details). A system of basis functions composed from M-modes is complete and orthogonal within a computational domain. The basis functions depend only on basin geometry and may be used many times for the same basin geometry.

The spectral entropy $S(K)$ may be applied to understand the complexity of spectrum a_k $(k = 1, \ldots, K)$ and the redistribution of SSH signals among different modes (Aubry et al., 1991). The entropy is calculated as

$$S(K, t) = -(\log K)^{-1} \sum_{k=1}^{K} p_k(t) \log p_k(t), \tag{3}$$

where $p_k(t) = a_k^2(t)/a(t)$, $a(t) = \sum_{k=1}^{K} a_k^2(t)$. $S = 1$ (0) for a uniform distribution of a signal among all K modes (the signal accumulated in only one mode).

Using the spectral entropy for the analysis of SSH variability has two explicit advantages. First, this simplifies the analysis of SSH variability because there is only one variable S instead of K variables a_k. Second, S is a measure to quantify temporal behavior of $\zeta(\mathbf{x}, t)$ and not only steady states.

The standard multi-fractal formalism (Muzzy et al., 1991; Pavlov and Anishenko, 2007; see also Appendix B) is used to understand and estimate how well the spectral entropy calculated by ROMS simulation results agrees with that calculated directly from observations. The key characteristic(s) estimated here are the scaling exponents $\theta(q)$ determined through the participation function (q-order structure function)

$$Z_q(\Delta t) = \frac{1}{T}\int_o^T |S(K, t+\Delta t) - S(K,t)|^q dt \sim \Delta t^{\theta(q)}, \quad (4)$$

where q is a real value parameter, $T \gg \Delta t$ is length of a ROMS run or observational series, and Δt^{-1} is sampling frequency. If $q = 2$, Z_2 is the classical structure function which is widely used in the statistical theory of turbulence (Monin and Yaglom, 1971).

The q-order structure function allows essential statistical features (correlations, scaling and non-linearity) in the behavior of entropy to be quantified, and thus provides understanding of the statistical similarity between surface elevations simulated by ROMS and extracted from satellite altimetry. The principle advantage of the approach is that the scaling exponents can be calculated for very small samples as well as when the observed signal is not stationary in the statistical sense.

3. Double spectral approach

To understand dominant spatial scales for SSH, $\zeta(\mathbf{x}, t)$ is decomposed using Eq. (2). The spectral density $E_\zeta(k, t)$ for SSH is defined as

$$\int\int \zeta(\mathbf{x}, t)\zeta(\mathbf{x}, t)d\mathbf{x} = \sum_{k=1}^{K} a_k^2(t), \quad (5)$$

$$a_k(t) = \int\int \zeta(\mathbf{x}, t)\psi_k(\mathbf{x})d\mathbf{x}, \quad E_\zeta(k, t) = a_k^2(t). \quad (6)$$

The basis functions ψ_k cannot mix contributions of different scales of a general circulation pattern as empirical orthogonal functions do, but modal decomposition (2) composed from these functions converges more slowly than similar ones for EOFs. However, since the basis functions ψ_k are quickly calculated because the matrix of the plane Laplace operator is sparse, and only groups of modes represent the physical meaningful current structures, the slower convergence of decomposition (2) is not a disadvantage for practical applications.

The dominant temporal scales of the flow are extracted from the spectral coefficients $a_k(t)$ using continuous or discrete wavelet transformations (Kumar and Foufoula-Georgiou, 1997). The former can be applied to specify energy-dominant scales but the latter is more convenient for filtering a multi-scale signal. It is always possible to select the coefficients with respect to the specified time scale and to reconstruct a filtered version of the flow from them, selecting only the energy-dominating scales (ridges on a scalogram). To perform this decomposition, a set of objective thresholds is used for selection. These thresholds are defined in terms of the spectral coefficients. Note the present study uses only a discrete wavelet transform because the inverse of a continuous wavelet transform may excessively smooth the reconstructed functions.

4. Data for ROMS-data comparison

4.1. RAFOS float data

Isobaric RAFOS float data were used in this study (Rossby et al., 1986). The ensemble of trajectories included 68 floats tracked during 1992–2007. The length of the float mission varied from 10 days to 1195 days, and a total of 26,054 float days were analyzed.

Float data may be visualized and downloaded at http://www.oc.nps.edu/npsRAFOS/. Technical details regarding the float observations are available in Margolina et al. (2006). Visual analysis of RAFOS float trajectories showed two types of westward drift. First, floats were be trapped by coherent mesoscale eddies, usually anti-cyclones with diameters less than 100 km, or coherent submesoscale vortices with diameters of 20–30 km, and subsequently transported southwestward within these eddies (Fig. 1a). If the speed of westward transport was estimated as a ratio of the distance between the initial and final positions of a float to the observational period, the mean drift velocity was equal to 2.0–3.0 cm/s (Garfield et al., 1999). These estimates were approximate because the drift velocity strongly varied across-shore and from season to season.

Second, other floats exhibited stochastic movements both nearshore and offshore (Fig. 1b). Float dispersion seemed to be similar to that observed in quasi-two-dimensional turbulent flows (Cardoso et al., 1996) where particle trajectories represent a combination of loops (“traps”) and streaks (“flights”) between loops. The zonal dispersion of floats in the mean was characterized by an effective diffusion coefficient of about 10^7 cm^2/s (Collins et al., 2004). The largest number of floats demonstrated this second type of westward drift.

4.2. Satellite altimetry

A SSH anomaly field using merged altimeter data sets has been produced by the AVISO Project (Archiving Validation and Interpretation on Satellite Data in Oceanography) for the period from October 10, 1992, to May 23, 2007. A typical SSH anomaly field extracted from the AVISO product is presented in Fig. 2. Details on the optimal interpolation and processing steps used to merge the different altimeter products (TOPEX/Poseidon, ERS-1/2, ENVISAT, Geosat Follow-On and Jason-1 altimeters) into a single map are given in Le Traon and Dibarboure (1999), Ducet et al. (2000) and Collecte Localisation Satellites (2006), and are not discussed here.

A goal is to find the spatio-temporal scales (if they exist) for which there is statistical similarity between the SSH anomaly field simulated by ROMS and those extracted from satellite altimetry. Preliminary calculations demonstrated that the dominant temporal scales were easily detected and selected in the SSH anomaly field using a discrete wavelet transform. As an example, an original SSH anomaly field (Fig. 3a) is decomposed into bi-annual and slower oscillations (Fig. 3b) and annual and faster oscillations (Fig. 3c). Note the quasi-zonal structures existing in Fig. 3b; the dynamics of these structures is studied in a separate paper (Ivanov et al., 2009).

4.3. ROMS configuration

The Regional Oceanic Modeling System (ROMS) has been developed by Shchepetkin and McWilliams (2003, 2005, 2008). The computational domain for the model extended in latitude from Cape Mendocino (about 41°N) to Pt. Conception (about 34°N) (Fig. 2). This domain included the most energetic eddy regions off Central California (Mackas, 2005). Horizontal grid spacing of the ROMS was 3.5 km; 20 vertical levels were used. This model was nested within another ROMS with 5 km horizontal grid

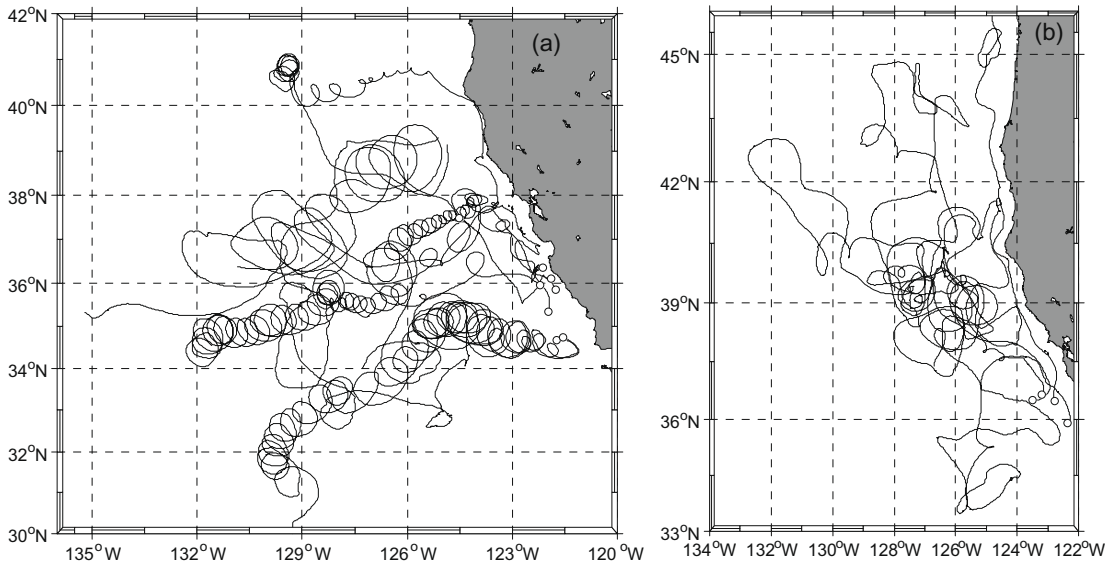

Fig. 1. Offshore transport shown by RAFOS floats launched between 300 m and 350 m. White dots are initial float positions. (a) Westward transport by coherent anti-cyclonic mesoscale eddies and submesoscale vortices, (b) turbulent-like float motions. The float trajectories in (b) are similar to those observed in 2D turbulent flow (Cardoso et al., 1996).

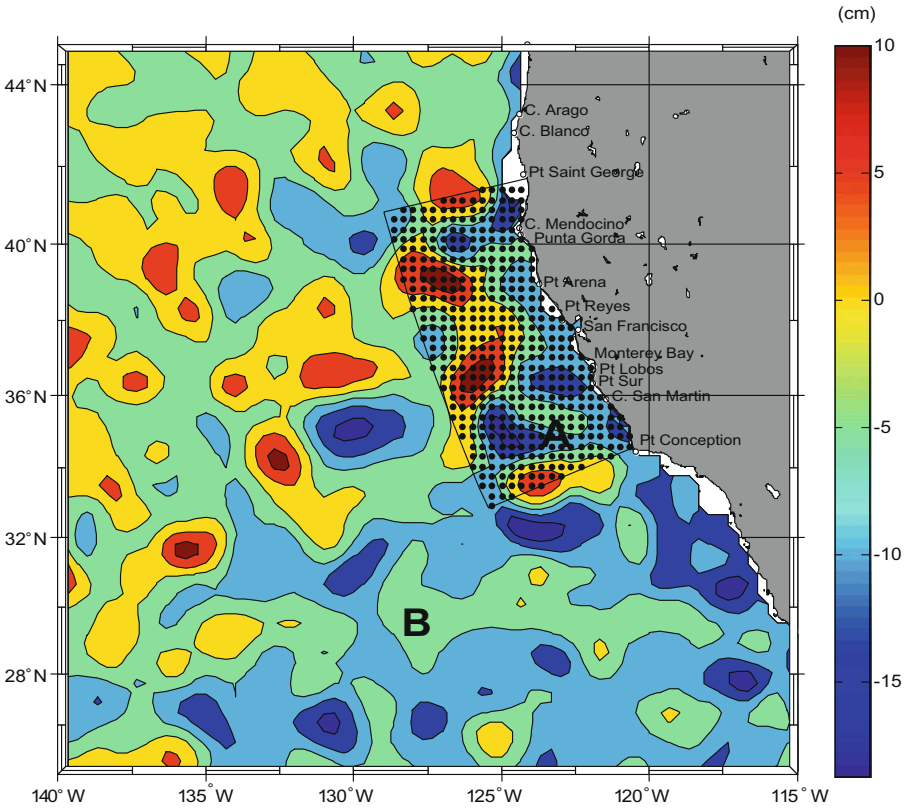

Fig. 2. A snapshot of the SSH anomaly field in the region off Central California (denoted by B) and the computational domain for ROMS (denoted by A). The SSH anomaly field was extracted from an AVISO product. Black dots show observational coverage for SSH.

spacing (Marchesiello et al., 2003). The model run used in the present study had a length of 5.5 years.

The model was forced by climatological seasonal wind stress, and heat and freshwater fluxes derived from the Comprehensive Ocean-Atmosphere data set (CODAS) taken from Da Silva et al. (1994). The model reached a spin-up state after 2 years of model integration when the surface kinetic energy oscillated quasi-periodically around an equilibrium state. The volume-integrated kinetic energy did not show a clear seasonal cycle, although it did exhibit inter-annual variability as an indication of intrinsic variability (Marchesiello et al., 2003). Marchesiello et al. (2003) has additional details of the ROMS configuration as well as oceanographic analysis of model results.

A typical SSH snapshot reproduced by ROMS is shown in Fig. 4. Note that although most large-scale coherent structures are clearly detected in the SSH, smaller eddy structures, which are easily

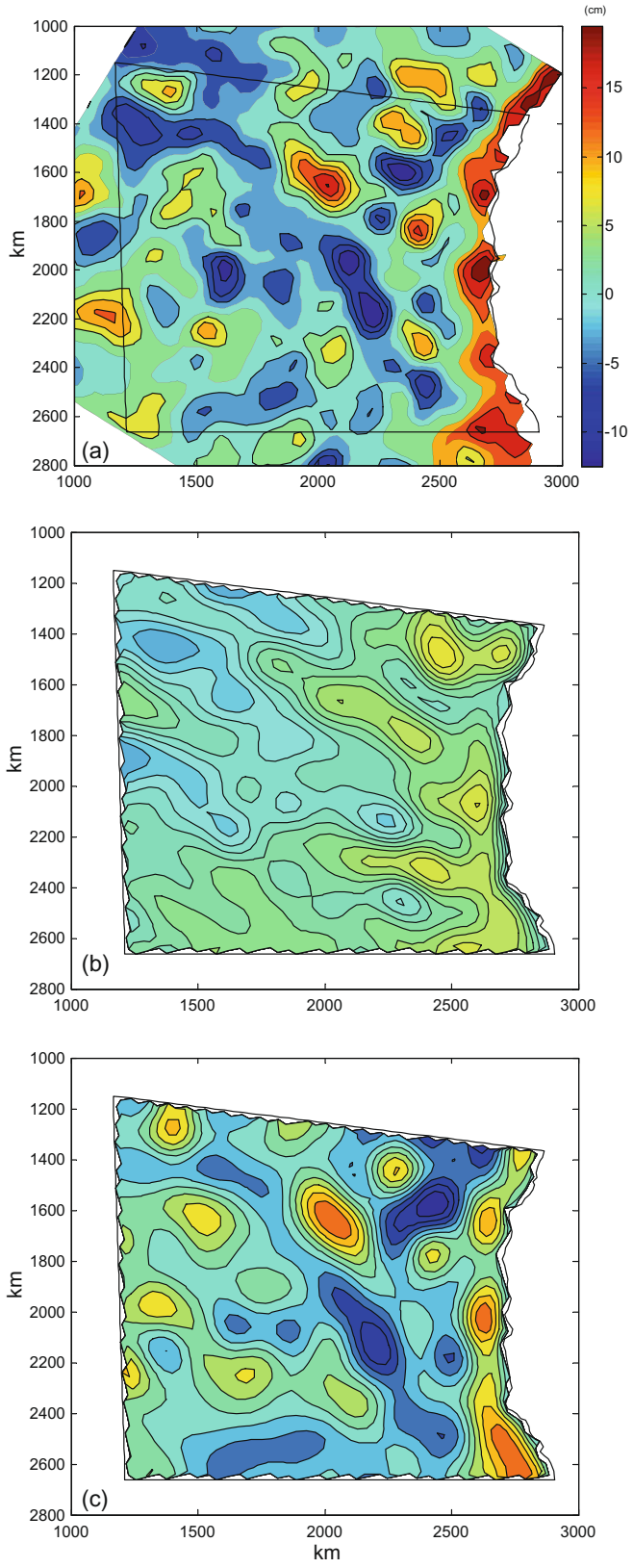

Fig. 3. Decomposition of the SSH anomaly field. (a) A snapshot of the SSH anomaly field, (b) SSH components with periodicity of 2 years and slower, and (c) SSH component with periodicity of 1 year and faster. Here and in Figs. 8–10 the computational domain was rotated clockwise 21°, and Cartesian coordinates were used instead of the geographic coordinates.

detected in circulation and temperature patterns, are masked in the SSH (Fig. 4). This effect was found earlier for an idealized submesoscale model of the California Current System (CCS) by Capet et al. (2008).

Since the computational grid step is equal to 3.5 km, the ROMS was able to reproduce submesoscale currents with spatial scales of 10–30 km. These currents consist of a chain of vortices along the coastline (as in Fig. 4) and single coherent vortices outside of mesoscale structures. Therefore submesoscale currents contributed considerably to westward transport off Central California within a 150–200 km zone, near the coastline, but not at large distances from the coastline.

5. ROMS–RAFOS float comparison

To compare ROMS results with RAFOS float observations, 1000 synthetic particles were launched in the ROMS simulated circulation field and tracked for 1 year. Particle trajectories were then used to estimate the westward transport and eddy diffusivity in the CCS. The same characteristics were calculated from the original RAFOS float data and compared to the synthetic particles using different techniques: a bin technique (Davis, 1991) and a technique for identification of float behavior through exit time statistics (see Artale et al., 1997; Ivanov et al., 2008, among others).

5.1. Bin technique

In principal, a flow can be represented as a mean circulation (usually assumed to be steady) and an eddy field (characterized by mean eddy kinetic energy), which are referred to centers of geographic bins. Mean velocity vectors and standard deviation ellipses computed from binned drifter velocities and high-resolution model simulations were compared to understand statistical similarity between the model and data (Garraffo et al., 2001; McClean et al., 2002, among others). Note that this approach is applicable only if the scale of the mean circulation is considerably larger than the energy-dominant scale of the eddy field (Corrsin, 1974).

This technique was applied to $0.5^{\circ} \times 0.5^{\circ}$ binned RAFOS float and synthetic particle velocity fields. Results for the ROMS and float data are shown in Fig. 5a and b, respectively. Comparison of these two figures leads to the following conclusion: although the California Undercurrent (a nearshore poleward flow) was clearly observed in both the fields, the mean current was spatially intermittent offshore. The floats indicated strong westward transport, but the mean circulation pattern for ROMS was more chaotic. Chaotic orientation of ROMS mean velocity vectors was observed in Fig. 5b and eddy variability was often comparable or larger than mean velocities.

However, the large difference between ROMS and RAFOS observations offshore (near 126°W) in Fig. 5a and b had a synthetic nature due to sampling errors of RAFOS float observations offshore and inaccuracy of the bin technique when velocities along all float trajectories are summarized within a bin, e.g. the geographic bins may on occasion split an eddy in half. No explicit gaps were found in the spatial spectrum of kinetic energy of mesoscale flow at mid-depth and accurate summation was needed to get the phase of Rossby wave-like perturbations traveling offshore (Ivanov and Collins, in press).

Using the traditional approach (Davis, 1991), the RAFOS float mean errors for estimates of mean velocities were 0.6 cm/s and 0.7 cm/s for zonal and meridional components, respectively. For ellipsoid axes, the mean error was about 0.4 cm/s. The corresponding errors estimated from ROMS were considerably smaller in part due to the larger sample size.

5.2. Exit time statistics

To quantify the transport processes in a more accurate manner for both the ROMS simulation and real float observations, the exit

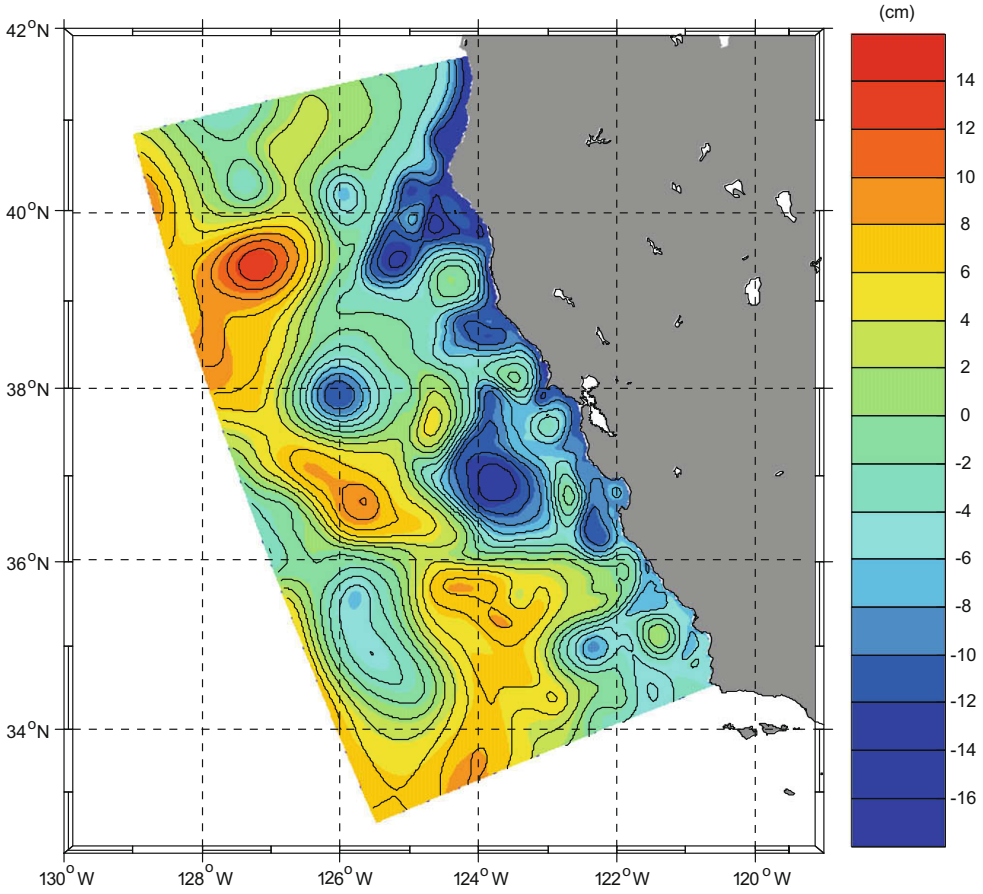

Fig. 4. Typical SSH field produced by ROMS in domain A (defined in Fig. 2). Non-dimensional black isolines correspond to the stream function. They indicate the spatial structure of the meso/submesoscale field.

time concept (Gardiner, 2004; Ivanov et al., 1994) and the random displacement model with the diffusion limit (Thomson, 1984) were applied, the latter without any assumption regarding the homogeneity of diffusion process. The velocity $\overline{u}(y)$ and effective across-shore diffusion coefficient $\overline{\chi}(y)$ averaged along-shore were introduced and used to characterize the westward transport simulated by ROMS and estimated from RAFOS float data. Hereafter, the overbar denotes along-shore averaging, and y is the distance from the coastline.

Since long-term correlations should accompany mesoscale flows reproduced by ROMS off Central California, parameterization of mesoscale transport as eddy diffusion seems to be not strictly justifiable. However, our previous experience (Ivanov et al., 2008) and results obtained by other researchers [see Pasquero (2005), as an example] have shown that if the effective diffusion coefficients are properly introduced, this parameterization is quite reasonable in many cases.

Using Lagrangian statistics calculated from the RAFOS float data for the CCS, Collins et al. (2004) demonstrated the existence of several correlation time scales for mesoscale oceanic flow off Central California: 9, 16 and 100–120 days. Therefore the displacement model should describe motions corresponding to first two scales and the last one as diffusion and advection, respectively, for time periods shorter than 100 days. If the period is considerably longer than 100 days, the displacement model should describe motion as diffusion.

To specify across-shore velocity $\overline{u}(y)$ and effective diffusivity $\overline{\chi}(y)$ from RAFOS float observations, M equidistant boundaries $(\partial\Omega_1,\ldots,\partial\Omega_M)$ were introduced, which were parallel to the smoothed coastline $\partial\Omega_o$ between 32°N and 47°N (Fig. 6a). The distance y_m from the mth boundary to the coastline satisfies a recurrent relationship: $y_m = m\Delta y$, where Δy is a spatial step. Then a set of exit times $(\bar{T}_m,\ m = 1,\ldots,M)$ was calculated, each of which had the length of time for a float to cross an appropriate boundary for the first time. Statistics of these exit times characterizes across-shore variability of the westward transport.

Since the number of floats was not large, in order to estimate moments of exit time as accurately as possible, the following approach was suggested. First, probability weighted moments (PWMs) (Greenwood et al., 1979) were used to reconstruct the probability density function of exit time and then to estimate the mean exit time. The moments were introduced as

$$\alpha_l(\Omega_m) = \int_0^1 X(P_m)(1-P_m)^l dP_m, \quad l = 1,\ldots,L, \tag{7}$$

where $X(P_m)$ is the quantile function, i.e. the inverse of cumulative distribution function for exit time $\bar{T}(\Omega_m)$, which is more robust relative to sampling errors and data outliers than convenient moments (Hosking and Wallis, 1997). In practice, the moments α_l were estimated from an ordered random sample $\{\bar{T}(\Omega_m)\}_{s=1}^S$ of size S (Hosking and Wallis, 1997) by

$$\hat{\alpha}_l(\Omega_m) = \frac{1}{S}(C_l^{S-l})^{-1}\sum_{s=1}^{S} C_l^{S-s}\{\bar{T}(\Omega_m)\}_s^S. \tag{8}$$

where C_l^s are the binomial coefficients.

Ivanov and Chu (2007a,b) suggested an iterative process to estimate T-PDF from small observational samples using PWMs and the

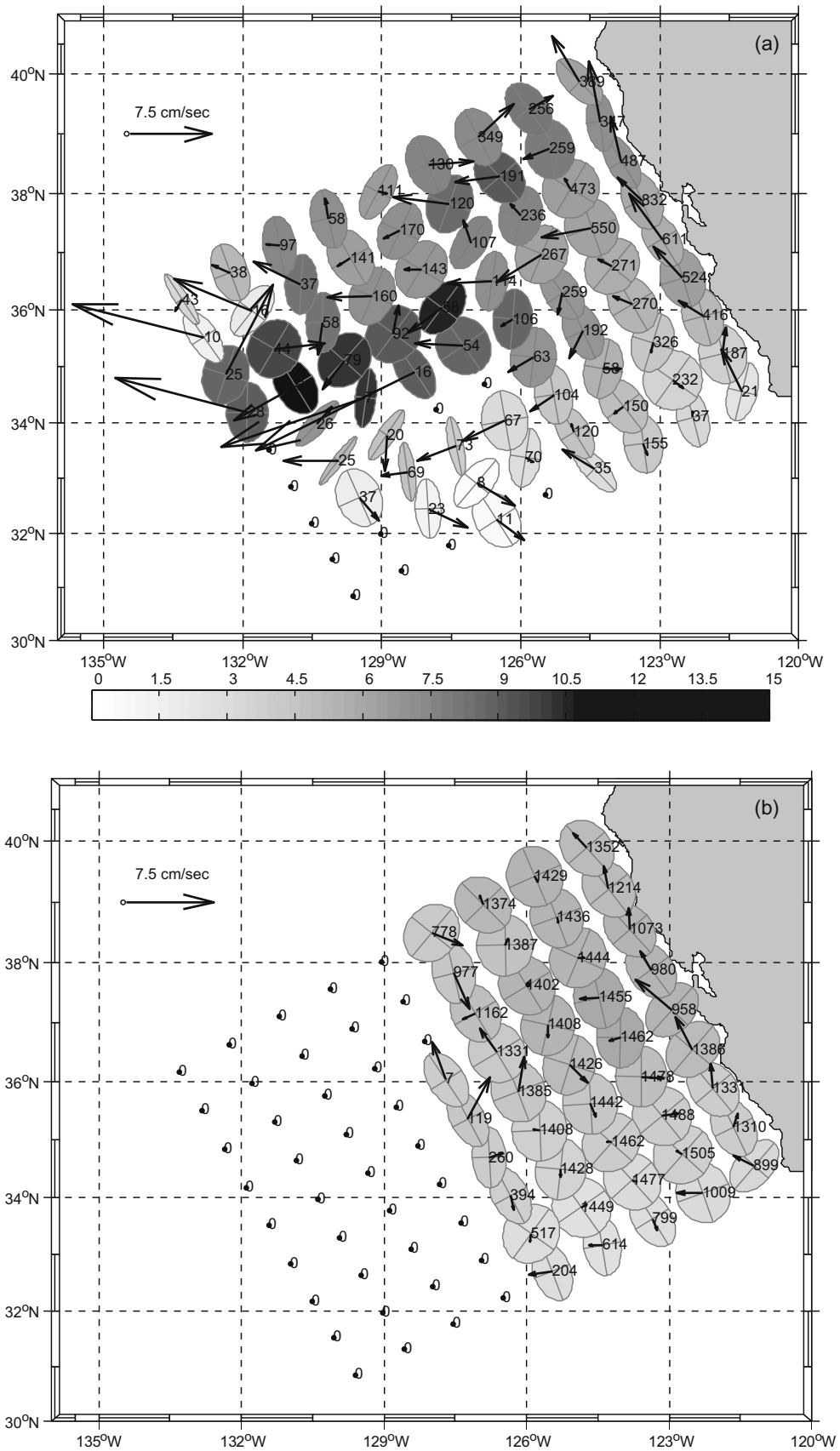

Fig. 5. Spatially binned ($0.5^\circ \times 0.5^\circ$) mean velocity and standard deviation ellipses. (a) RAFOS float observations. (b) Synthetic particles deployed in the ROMS simulated circulation. Velocity scale is indicated in the upper left hand corner. The total number of floats or particle observations for each box is indicated in small fonts.

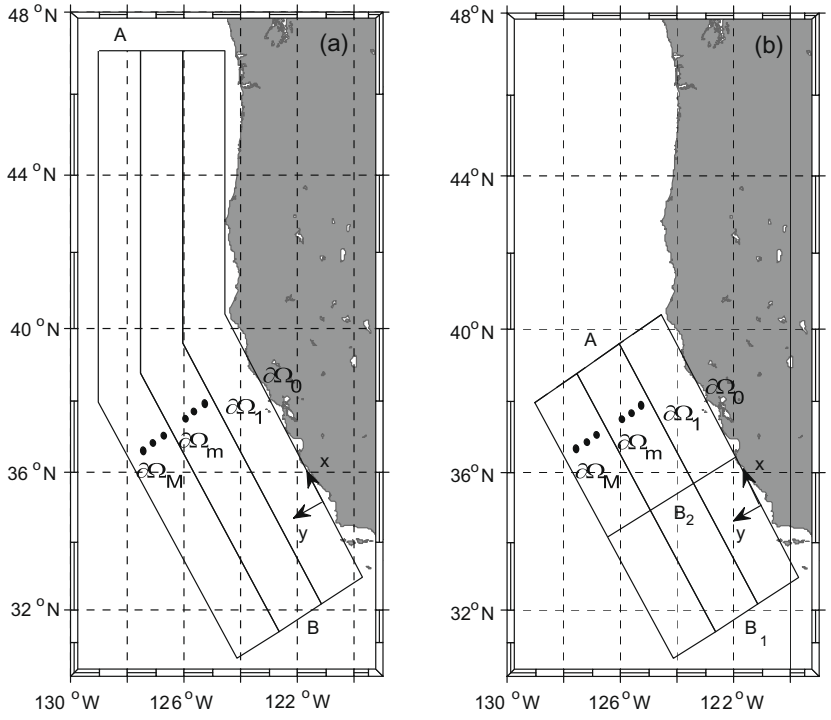

Fig. 6. Sub-domains for exit time analysis. (a) Ω_m ($m = 1,\ldots,M$). Ω_m are bounded by $\partial\Omega_m \cup \partial\Omega_o \cup A \cup B$. (b) Two sets of reduced sub-domains used for sensitivity analysis Ω'_m ($m = 1,\ldots,M$) and Ω''_m ($m = 1,\ldots,M$). Ω'_m and Ω''_m are bounded by $\partial\Omega_m \cup \partial\Omega_o \cup A \cup B_1$ and $\partial\Omega_m \cup \partial\Omega_o \cup A \cup B_2$, respectively. x- and y-axes are along and across-shore directions, respectively.

maximum entropy principle. Numerical details of the procedure were discussed in the aforementioned papers. It was noted that this procedure can be applied to observational sample sizes as small as 10–20, and a three-parametrical Weibull distribution was used as the first guess for the iterative process. This approach sometimes specified the mean exit times $\langle \bar{T}_m \rangle$ in a very accurate manner even if only 10 float trajectories were used.

Second, to estimate the variance of exit time from float data, additional information regarding variance is required because its estimate is more sensitive to length of observational sampling than the mean exit time. Following Ivanov et al. (2008), it was *suggested* that the mean exit time averaged along-shore grew as a power law

$$\langle \bar{T} \rangle = \mu y^{\beta} + \gamma, \tag{9}$$

where y is the distance from the coastline (Fig. 5a) and β, μ and γ are parameters. Eq. (9) is the best fit to results obtained from RAFOS float observations. The same data were used in this study to compare ROMS to the RAFOS data.

The power law assumption for the mean exit time (9) implies a similar dependence for the drift and diffusivity $\overline{v}(y) = v_o y^{1-\beta}$ (cm/s), $\overline{\chi}(y) = \chi_o y^{2-\beta}$ (cm^2/s), where $\overline{v}(y)/\overline{\chi}(y) = \gamma_o y^{-1/2}$ (cm^{-1}), and $\gamma_o = 2v_o/\chi_o$ is a non-dimensional and positive parameter. Here, $\overline{v}(y) = \overline{u}(y) + \frac{1}{2}\nabla\overline{\chi}(y)$ is float velocity, ∇ is the across-shore gradient operator, and $\nabla\chi$ is additional drift induced by inhomogeneity of the mesoscale field (Monin and Yaglom, 1971).

Ivanov et al. (2008) analytically found that if condition (9) holds, then

$$\langle \bar{T} \rangle = \mu(y^{\beta} - \langle \sigma^{\beta} \rangle), \tag{10}$$

$$\langle \delta\bar{T}^2 \rangle = \frac{\mu^2}{(\gamma_o + 2\beta - 1)}(y^{2\beta} - \langle \sigma^{2\beta} \rangle), \tag{11}$$

where $\mu = \frac{1}{\chi_o \beta(\gamma_o+\beta-1)}$, σ is the launch position of a float along the y-axis, $\langle \ldots \rangle$ is averaging over the initial distribution of floats $f_o(\sigma)$ (determined from the RAFOS observations or initial distribution of synthetic particles).

Using (10) and (11), the parameters β, μ and γ were estimated in a two-step procedure. For the first step, $\langle \bar{T} \rangle$ was smoothed to $\langle \bar{T}_m \rangle$ in the least squares sense to obtain the first guess for v_o^*, χ_o^* and β^*. For the second step, parameters v_o, χ_o and β were estimated from the following variational problem

$$\min_{\beta, v_o, \chi_o} |\mathbf{R}|^2, \quad \min_{\beta, v_o, \chi_o} \int_o^y [\langle \delta\bar{T}_m^2(\xi) \rangle - \langle \delta\bar{T}^2(\xi) \rangle] d\xi, \tag{12}$$

where $\langle \delta\bar{T}_m^2 \rangle$ is the variance computed directly from float data by the Monte-Carlo sampling technique (Good, 1996), $\left(\frac{\partial\langle\delta\bar{T}^2\rangle}{\partial\beta}, \frac{\partial\langle\delta\bar{T}^2\rangle}{\partial v_o}, \frac{\partial\langle\delta\bar{T}^2\rangle}{\partial\chi_o}\right)$ is the transposed sensitivity vector $\mathbf{R}^{tr}$. The variation problem (12) was solved through the iteration procedure developed by Chu et al. (2004b) where the first guess was v_o^*, χ_o^* and β^*.

The methods discussed above permit robust estimates of the power exponent β (control parameter) and parameters v_o and χ_o from a small number of observations. Neither the mean exit time nor its variance were sensitive to perturbations of parameters χ_o and v_o.

Theoretical model (10) and (11) allow diffusion to be distinguished from advection. If $\chi_o = 0$ (no diffusion), then $\langle \delta\bar{T}^2 \rangle = 0$, and float motion is an advective process. In the opposite case, when $\bar{u} = 0$ (no advection), $\gamma_o = \frac{1}{2\chi_o}\nabla\bar{\chi}$, and $\mu = \frac{1}{\chi_o\beta}$, i.e. float motion is a diffusive process characterized by two parameters, χ_o and β. For homogenous diffusion with a constant diffusion coefficient χ_o, $\beta = 2$ and $\mu = \frac{1}{2\chi_o}$. Therefore, homogeneous diffusion is distinguished from inhomogeneous diffusion by the value of μ, and an advective process distinguished from a diffusion process by the behavior of $\bar{T}$ – variance with y. If advective processes dominate float motion between boundaries $\partial\Omega_{m-1}$ and $\partial\Omega_m$, the value of $\bar{T}$ – variance calculated for sub-domain Ω_m is the same as for Ω_{m-1}.

5.3. Exit time statistics for ROMS and RAFOS floats

The mean exit time averaged along-shore from Cape Mendocino to Pt. Sur is shown in Fig. 7a. Results demonstrated good agreement between the ROMS simulated particles and the RAFOS floats as far as 350 km from the coastline. In the mean, the particles and floats move offshore in the same manner, and with similar kinematic characteristics. The mean cross-shore transport was characterized by variable diffusion

$$\overline{\chi}(y) = \chi_o y, \tag{13}$$

where $\chi_o \approx 0.7-0.8$ cm/s and $\beta \approx 1$.

The choice of β strongly affected the behavior of the variance of exit time (a deviation greater than 10% of β from the optimal value $\beta \approx 1$ resulted in large variations of variance). Differences between power exponents β estimated from the ROMS simulation and RAFOS observations did not exceed 5% for $y \leqslant 350$ km. These differences were less than the confidence interval for estimates of β obtained by Ivanov et al. (2008) directly from RAFOS float data.

The estimates of the diffusion coefficient are less sensitive to the choice of parameters v_o and χ_o. Therefore, a difference as large as 25% between these parameters estimated from the ROMS simulated particles and float data did not seem to be crucial.

For $y > 350$ km, ROMS overestimated the speed of across-shore transport by a factor of approximately 1.2–1.3 as compared to RAFOS float observations (Fig. 7a). Two reasons explain the greater across-shore transport for ROMS. First, estimates of kinematic characteristics of the across-shore transport from RAFOS float data were less accurate offshore because only a portion of the floats travelled this far from the coast. Second, accuracy of ROMS hindcast near open boundaries was reduced due to limited domain size and use of climatological conditions at the open boundary.

The variance of exit time shown in Fig. 7b also indicated that coherent structures dominated the across-shore transport in ROMS and were statistically similar to those responsible for the along-shore-averaged westward transport of the RAFOS floats. Note that within the region between 180 km and 250 km, both ROMS and float observations indicated the existence of a coastal transition zone (CTZ) that separated coastal and deep ocean waters and within which the westward transport is due to filamentary jets, i.e. the transport was advective by nature [see Brink et al. (2000) among others where the CTZ was identified in the upper ocean using other observations]. *In-situ* and satellite observations have shown that the westward transport across the CTZ is caused by filaments and seems to be an advective process (Miller et al., 1999). Our results agreed with this conclusion.

For $y > 250$ km, ROMS when compared to RAFOS float data underestimated the variability of across-shore transport because the standard deviation of the exit time estimated from the float data was 1.5–1.8 times larger than that calculated from ROMS. This difference was probably due to the idealized external forcing used in ROMS.

The mean exit time, in principle, can be parameterized differently than (9). However, no reasonable estimates for both the mean exit time and its variance were found to be better than those expressed by the power law. It was not possible to improve the estimate of the mean exit time without degrading the accuracy of the estimate of the variance and vice versa.

These results were sensitive neither to reduction of the length of the along-shore dimension of the computational domain nor to variations of Δy between 5 km and 50 km. Two sets of smaller sub-domains, Ω'_m and Ω''_m, shown in Fig. 6b were used to make sure that reduction of computational domain size Δy did not result in a new dynamical regime of float dispersion or different values of β.

6. ROMS-satellite altimetry comparison

As noted above, comparison of ROMS with satellite altimetry is possible only for specific spatio-temporal scales. If it is hypothesized that temporal variability of mesoscale circulation in the

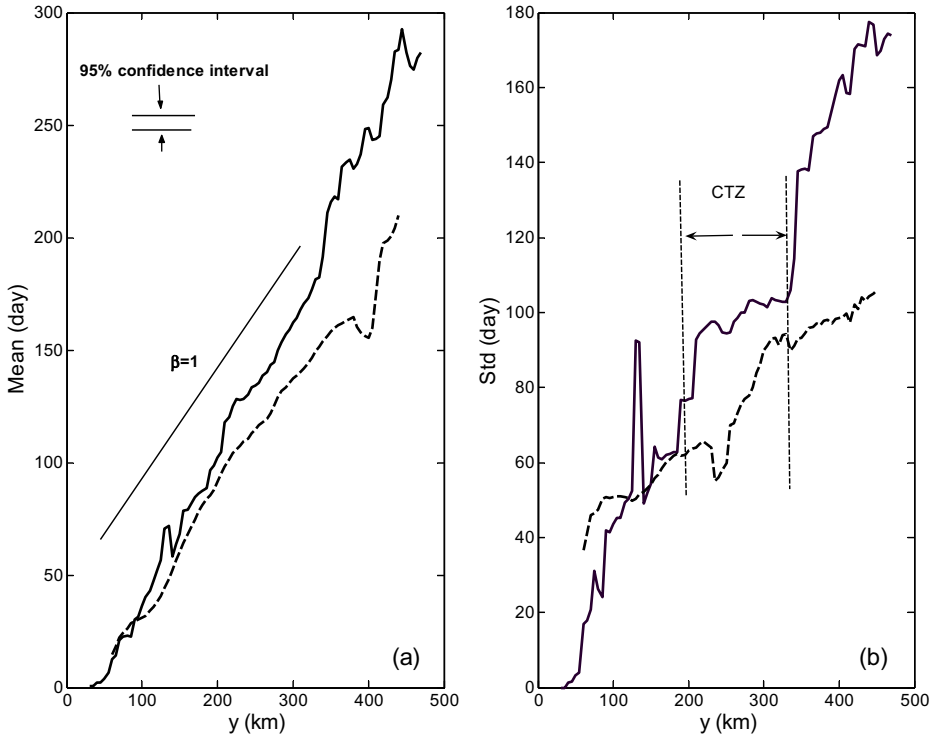

Fig. 7. Exit time statistics. Solid and dotted curves are estimates for RAFOS floats and synthetic particles deployed in ROMS simulations, respectively. (a) Mean exit time. Black line is scaled law with a power exponent equal to 1. (b) Standard deviation of exit time. The location of the coastal transition zone (CTZ) is indicated by the vertical dashed lines.

CCS is likely to be affected by inter-annual variability, ROMS and satellite altimetry may be compared for the temporal scales of 1 year and faster because ROMS was forced only by seasonal winds (Marchesiello et al., 2003).

On the other hand, variability of the SSH anomaly field extracted from the AVISO product cannot be faster than about 1 month due to observational constraints. Although the SSH anomaly field gridded to a $0.25^\circ \times 0.25^\circ$. Mercator grid was represented weekly, each time-point incorporated 6 weeks of observations weighted to the central date and merged over multiple satellites to reduce spatial error (Le Traon et al., 2003). Therefore, the real temporal resolution is determined by the frequency of satellite observations, which was no better than $O\,(1/20\,\mathrm{day}^{-1})$.

It seems reasonable to compare ROMS to satellite altimetry for temporal scales ranging from 1 month to 6 months. The technique discussed in Section 3 was used to detect and select such energy-dominant scales in the SSH anomaly field.

6.1. Calculation of spatial modes

Basis functions $\{\psi_k\}_{k=1}^{1000}$ were calculated on an adaptive triangular grid in the computational domain Ω (Fig. 8). Using this grid allows (a) a smaller number of computational grid knots (66,414) than the original grid (71,440) used by Marchesiello et al. (2003), and (b) more accurate representation of nearshore circulation because the smallest triangles were used close to the coastline. The latter improved convergence of modal decomposition (2). Spectral problems for the plane Laplace operator were numerically solved using the MATLAB PDE toolbox (www.mathwork.com/r2007a). A few basis functions (low-order, intermediate and high-order) are shown in Fig. 9.

Note that these modes were easily computed, and required neither a large amount of computer time nor a high-performance computer. Our estimate showed that less than 200 modes were required to reproduce the SSH anomaly field with reasonable accuracy. As an example, an SSH snapshot reproduced by ROMS (Fig. 10a) and its counterpart represented by 200 modes (Fig. 10b) were very close to one another. RMSE between the original and reconstructed fields was less than 2–3% (for points where SSH was measured) and their residual was less than 0.01 cm.

Following Aubry et al. (1991), an estimate, K_o, of how many of the largest coefficients should be retained to give an efficient, low-entropy, representation of SSH $\zeta(\mathbf{x}, t)$ is given by

$$K_o = \lim_{K\to\infty} K^{S(K)}. \tag{14}$$

Variations of K_o for the SSH anomaly field reproduced by ROMS and that extracted from the AVISO product are shown in Fig. 11a and b, respectively. Here, K_o was estimated for the different number of modes in decomposition (2) to determine the convergence of (14) as $K \to \infty$. Independently of season and year, $K_o < 35$ and $K_o < 80$ for satellite altimetry and ROMS, respectively. This indicates an accumulation of SSH signal in the lowest order modes.

ROMS evidently has more degrees of freedom than SSH observations due to the limited space-time resolution of satellite altimetry. This additionally supports the above conclusion that ROMS results and altimetry data can be compared only within a specified time-scale band.

6.2. Selection of spatio-temporal scales

SSH produced by ROMS was represented as

$$\zeta(\mathbf{x}, t) = \zeta_o(\mathbf{x}) + \zeta_1(\mathbf{x}, t) + \zeta_2(\mathbf{x}, t) + \zeta_3(\mathbf{x}, t), \tag{15}$$

where ζ_o, ζ_1, ζ_2 and ζ_3 were steady (mean) circulation, low-frequency (oscillations with periodicity slower than 1 year), middle-frequency (oscillations with periodicity faster than 1 year but slower than 1 month) and very-high frequency (oscillations with periodicity faster than 1 month) currents, respectively.

As described above, ROMS-altimetry comparison can be possible only for temporal scales from 1 month to 6 months. Therefore, seasonal, inter-annual and very high-frequency variability were excluded from the analysis. To select appropriate scales, SSH anomaly was represented in form (4) and a Daubechies wavelet transform of 5th order with 6 levels (Daubechies, 1992) was used to filter appropriate motions from the total altimetry signal. Detailed explanations on how to use discrete wavelet transforms for signal

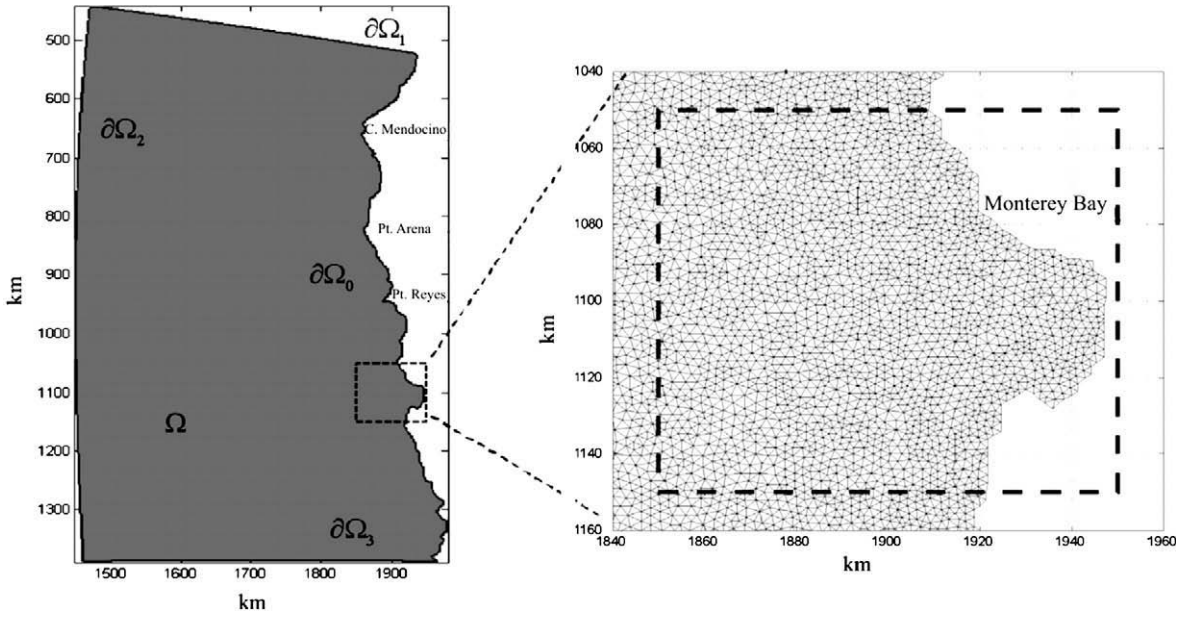

Fig. 8. Domain Ω for calculation of the basis functions ψ_k. A triangular adaptive grid was used for computation of basis functions; an example of the grid near Monterey Bay is shown in the inset.

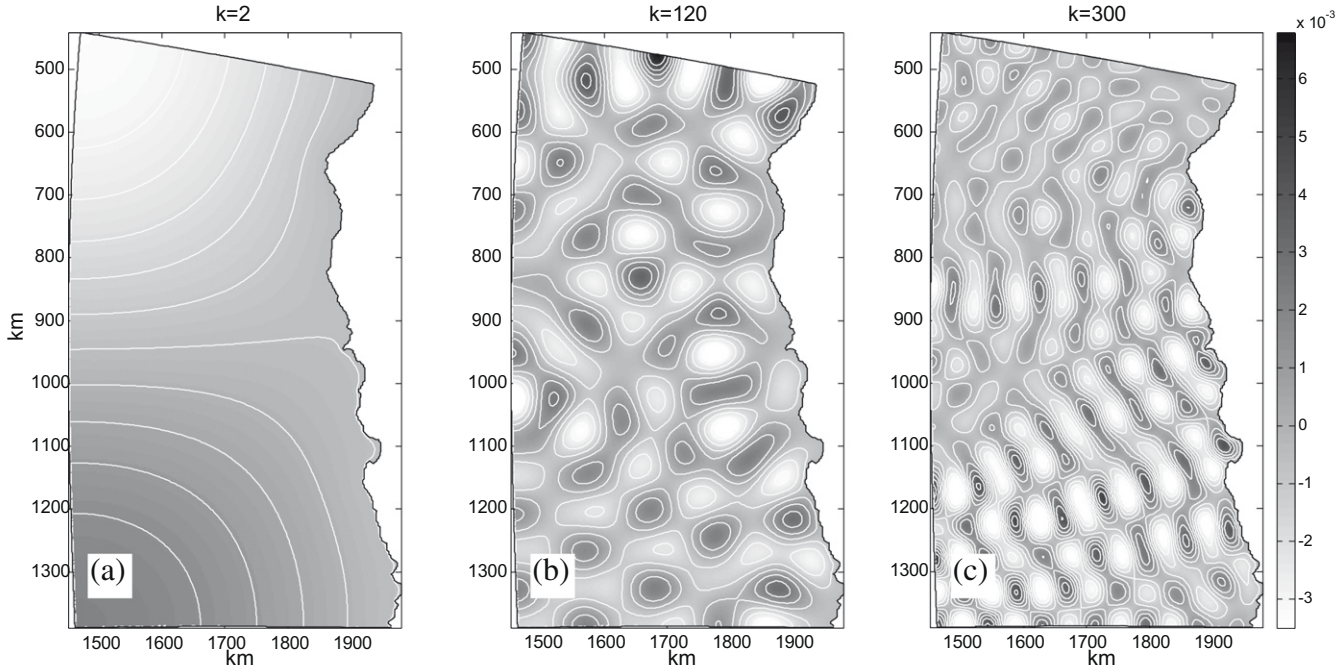

Fig. 9. Spatial distributions of ψ_k modes. (a) $k = 2$, (b) $k = 120$ and (c) $k = 300$. The contour interval is non-dimensional, with positive vorticity in dark and negative vorticity in light.

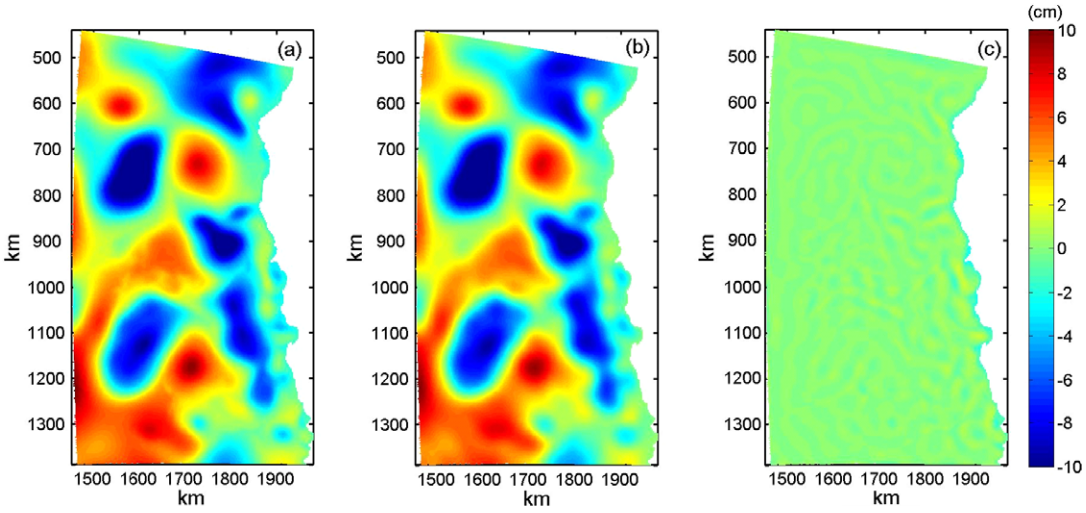

Fig. 10. Modal representation of SSH anomaly field reproduced by ROMS. (a) SSH anomaly field. (b) Two-hundred mode decomposition of the SSH anomaly field. (c) Residual field [difference between (a) and (b)]. The non-dimensional isolines indicate the structure of mesoscale field.

filtering are given by Kumar and Foufoula-Georgiou (1997). The capability of this approach is illustrated by Fig. 12a–c where different temporal components of the SSH anomaly signal extracted from AVISO products and averaged over 15 years are shown. Fig. 12a shows that inter-annual oscillations (dashed curve) are comparable to faster oscillations (solid curve) only for the largest spatial scales corresponding to modes numbered from 1 to 4. It is reasonable to hypothesize that these oscillations are excited by large-scale components of wind which may pump energy for the most part into low-order modes [see for example, Ivanov and Chu (2007b)].

The spectrum of signals with annual periodicity and slower is shown in Fig. 12b (dashed curve). This spectrum decayed in an exponential manner and discrete peaks were clear at k =1, 4, 6, 8, 11, 16, 19, 22 and 29. These oscillations dominated the altimetry signal for all modes lower than K_o (denoted the by black arrow in Fig. 12b). Oscillations with semi-annual periodicity and faster did not dominate the altimetry signal for lowest order modes, but were comparable to annual oscillations for low-order modes $k \geqslant 14$, and even became more intense than the annual oscillations for $k > 30$. High-frequency (1 month and faster) signals (Fig. 12c, solid curve) looked similar to white noise in the spectral band between $k = 2$ and $k = 30$ where the spectrum shape was nearly flat. Therefore, these signals were not analyzed.

Filtered spectral coefficients $\hat{a}_k$ that corresponded to the ζ_2 component of SSH were used to calculate time-averaged SSH spectra (here, averaging was done over length of the ROMS run or over the length of the SSH observational series) and spectral entropy. These results allow estimates of contributions from different temporal scales to the total SSH signal as well as the distribution of these contributions among different spatial modes.

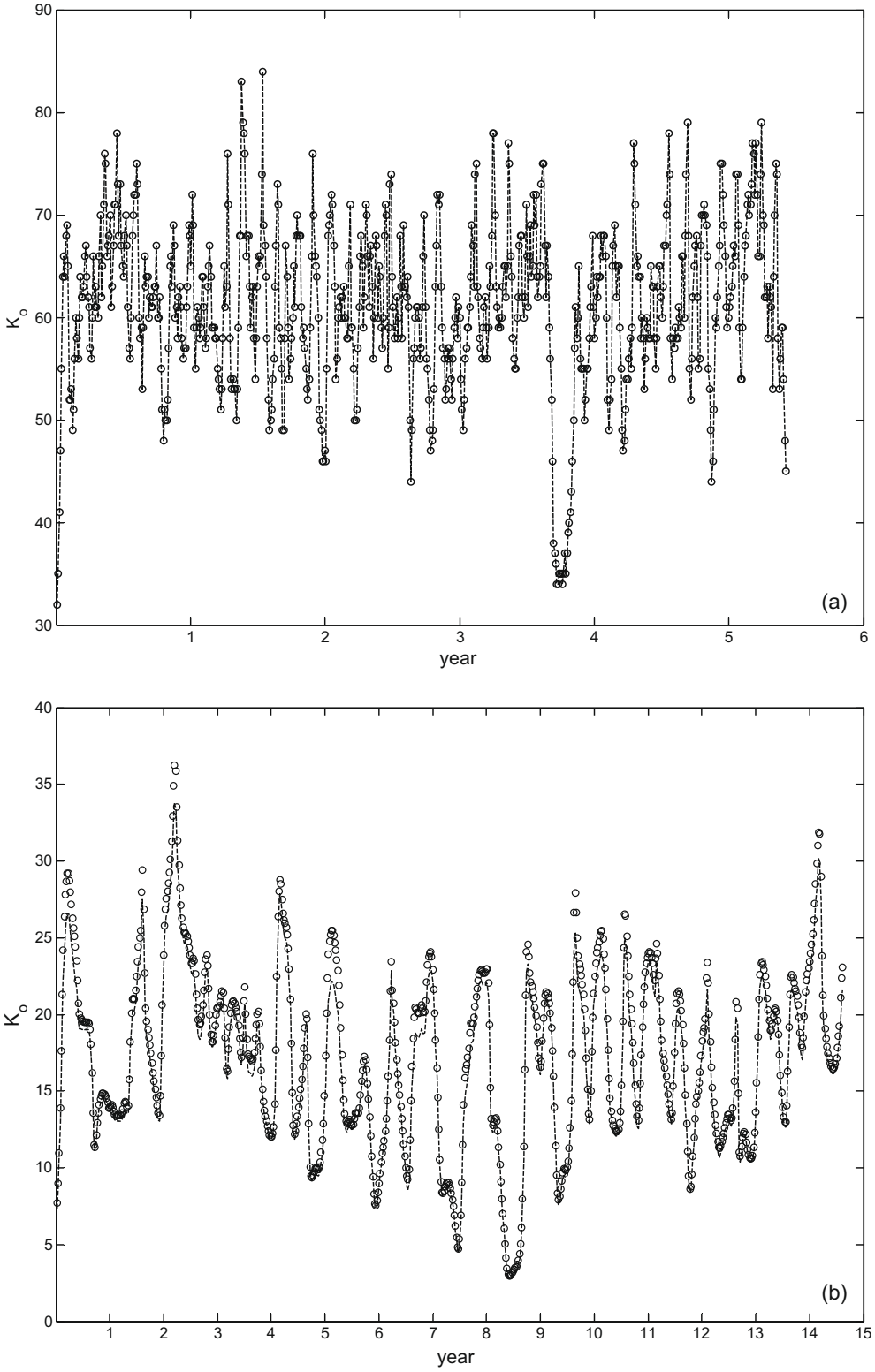

Fig. 11. Number of dominant modes K_o in modal decomposition (2) calculated for (a) ROMS simulation and (b) satellite observations. White circles and black curves correspond to $K = 200$ and $K = 100$ modes, respectively.

Fig. 13 shows the time-averaged SSH anomaly spectra calculated from the ROMS simulation results (solid curve) and AVISO data (dashed curve) for temporal scales of 1–12 months. These calculations indicated that (a) both mean spectra were similar for $2 \leqslant k \leqslant 60$, i.e. within a spectral band δk containing 59 modes and which corresponded to the dominant scales of SSH anomaly signals (K_o is denoted by a solid black arrow in Fig. 13). A difference between the low-order modes is explained by the climatological boundary conditions and forcing of ROMS; (b) both spectra showed that the same scale, $K_p \sim O(100\,\mathrm{km})$, contained the most energy (Fig. 13) and (c) the spectra were considerably different for higher-order modes ($k > 60$). The

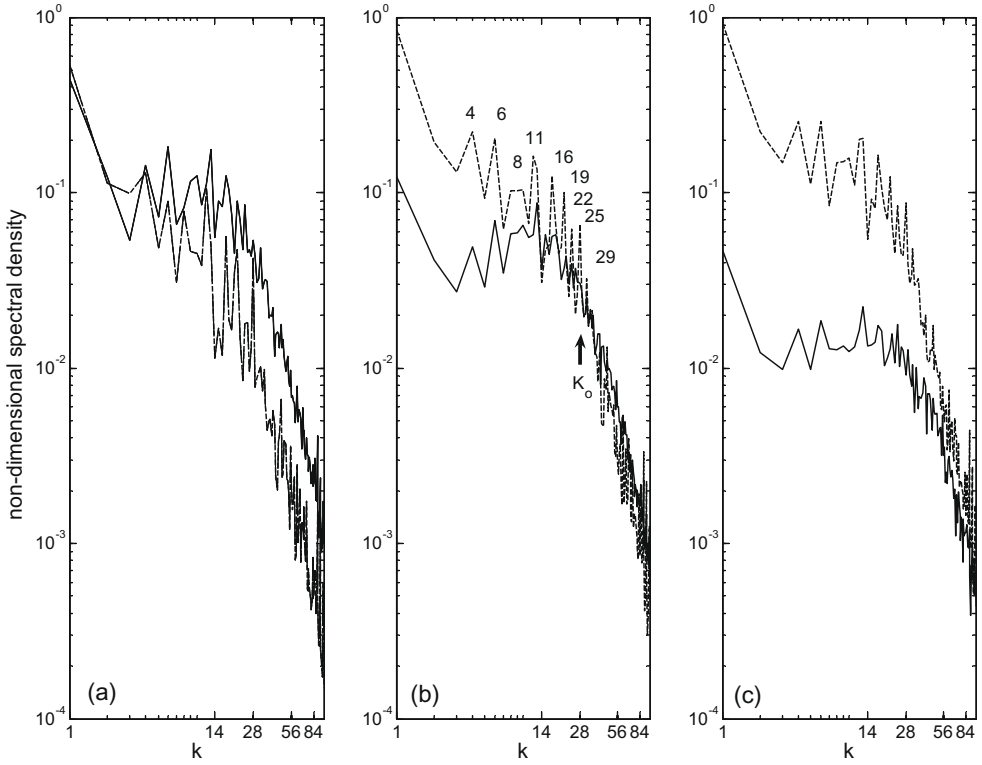

Fig. 12. Time-averaged spatial spectra of different temporal components of the SSH anomaly field calculated from satellite altimetry and normalized by time-averaged E_ζ. (a) For oscillations slower than 1 year (dashed curve) and oscillations of annual and faster periodicities (solid curve). (b) For oscillations of annual and lower periodicities (dashed curve) and oscillations of semi-annual and faster periodicities (solid curve). Black arrow indicates the wave number K_o. Discrete sharp peaks are numbered from 4 to 29. (c) For oscillations of 1 month and slower periodicities (dotted curve) and for oscillations of periodicity faster than 1 month (solid curve).

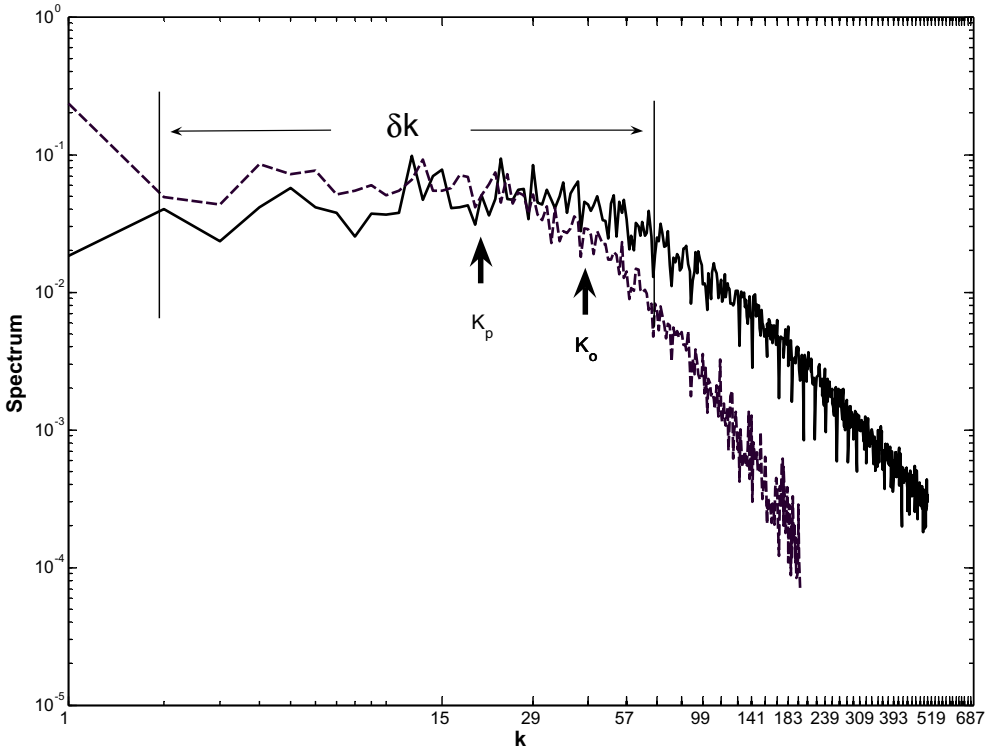

Fig. 13. Comparison of time-averaged SSH spectra for a time scale range of 1 month to 12 months. Solid and dashed curves are from the ROMS simulation and satellite data, respectively. Black arrows indicate the wave numbers K_o and K_p.

spectrum for the SSH anomaly field extracted from the AVISO product decayed in a faster manner than the one obtained for the ROMS simulations. This indicates that the SSH anomaly field was smoothed at these shorter scales to reduce spatial errors caused by the inaccuracy of the AVISO product in coastal domains.

6.3. Spectral entropy

Spectral entropies calculated from the ROMS simulation results and from SSH anomaly fields for a range of temporal scales of 1–12 months are shown in Fig. 14a and b. Although mean values (variances) of the spectral entropy for these time series were

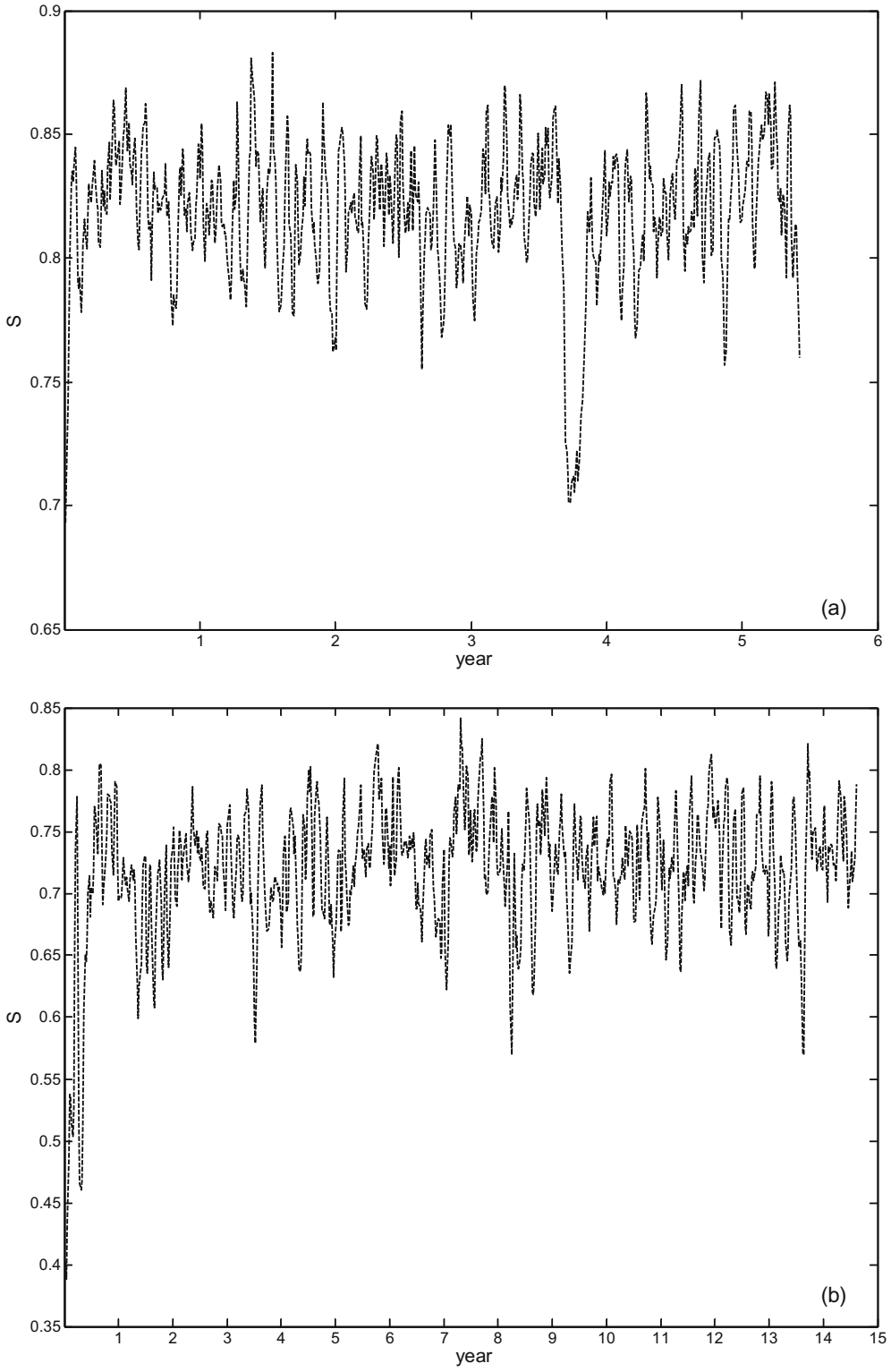

Fig. 14. Spectral entropy calculated for a time scale range of 1 month to 12 months from (a) ROMS simulations, and (b) satellite altimetry data.

0.79 (0.01) and 0.68 (0.02), respectively, it is hard to understand whether these entropies were statistically similar.

In Section 2 it was pointed out that highly variable signals such as the spectral entropy for the sea level elevation reproduced by ROMS and SSH derived from satellite altimetry can be compared using the appropriate q-order structure function. From a computational perspective, it is more convenient if the structure functions and $\theta(q)$ are not calculated directly but instead the singular exponent $\tau(q)$ is estimated through a wavelet transform technique. The following relationship between the scaling exponent of the structure function and the scaling singular exponent $\tau(q)$ exists (Muzzy et al., 1991):

$$\theta(q) = 1 + \tau(q). \tag{16}$$

By using the Legendre transformation (see Appendix B for explanations), the multi-fractal spectrum $D(h)$ can be obtained from $\tau(q)$ where $h = d\tau/dq$ is Hölder exponent. $D(h)$ captures how "frequently" a value h is found (Pavlov and Anishenko, 2007).

A multi-fractal formalism based on wavelet transforms [the wavelet transform modulus maxima method (WTMM)] was applied to both spectral entropies to estimate $\tau(q)$ and then to understand their differences and similarities. Inspired by the early work of Mallatt (1989) on the use of wavelet projections to provide a multi-resolution representation of signals, the concept of WTMM was originally introduced by Mallatt and Zhong (1992) in the context of signal processing and later applied by Muzzy et al. (1991) to the analysis of turbulent flows. Details of this approach can be found in Appendix B.

Singular exponents $\tau(q)$ and multi-fractal spectra $D(h)$ are shown in Fig. 15a and b. Using the same sliding windows for the ROMS and satellite altimetry, $S-1$ sub-samples were obtained, and then singular exponents and multi-fractal spectra were calculated for each of them. Then, results of averaging over the ensemble of sub-samples allow estimates of the mathematical expectation and confidence intervals for these characteristics. This is, in essence, one of the re-sampling methods (Good, 1996).

Comparing results for ROMS and satellite altimetry leads to the following conclusions:

(a) If the similarity of the spectral entropies calculated from ROMS and satellite altimetry was estimated by scaling exponents $\tau(q)$, then both the entropies seem to be very close for $-8 < q < 4$ (Fig. 15a). Differences between scaling exponents for both entropies may be explained by larger errors in estimating high-order structure functions due to sampling errors. The values $q > 0$ ($q < 0$) corresponded to large (small) fluctuations of the spectral entropy.

(b) There was a shift of $\Delta h = 0.15$ between multi-fractal spectra (Fig. 15b). However, for both the entropies $D(h)$ reached its maximum value as $h > 0.5$: 0.75 and 0.9 for satellite data and ROMS results, respectively. This was a manifestation of large correlations existing in the entropy behavior: the probability for a large value of entropy following a small value is considerably less than the probability for a large (small) value following a large (small) value (Pavlov and Anishenko, 2007). Difference between widths of the spectra (the width of the spectrum characterizes the strength of multi-fractal effects) is 0.03–0.05. This seems a small difference.

7. Conclusions

The present study compared ROMS simulations with a horizontal grid spacing of 3.5 km with two different data sets (RAFOS floats and satellite altimetry) to understand the ability of the model to reproduce meso/submesoscale currents for the Pacific Ocean off Central California. In contrast to the traditional oceanographic comparisons, this paper introduced new metrics for model-data comparison and suggested techniques to calculate them from short observational series which are non-stationary in the statistical sense. An important detail of this approach is that the double spectral analysis (Ivanov and Collins, in press) was used so that dominant scales could be detected and selected in complex SSH signals produced by ROMS and the satellite altimetry data (AVISO product). No explicit gaps to specify Nyquist "frequency" were found in the spatial spectrum of SSH anomaly field but it was possible to separate temporal scales.

The principal conclusion is that ROMS with sufficient resolution reproduced the main statistical properties of the mesoscale circulation field at surface and mid-depth (350 m). Specifically, ROMS reproduced in an accurate manner the kinematic characteristics of mean westward transport observed in RAFOS float motions at mid-depth for distances to 350 km from the coastline. For distances larger than 350 km, the ROMS overestimated the speed of westward transport and underestimated its variability. This difference between ROMS and RAFOS float data was likely due to the idealized (climate) external forcing for ROMS and poor statistics for the few floats observed.

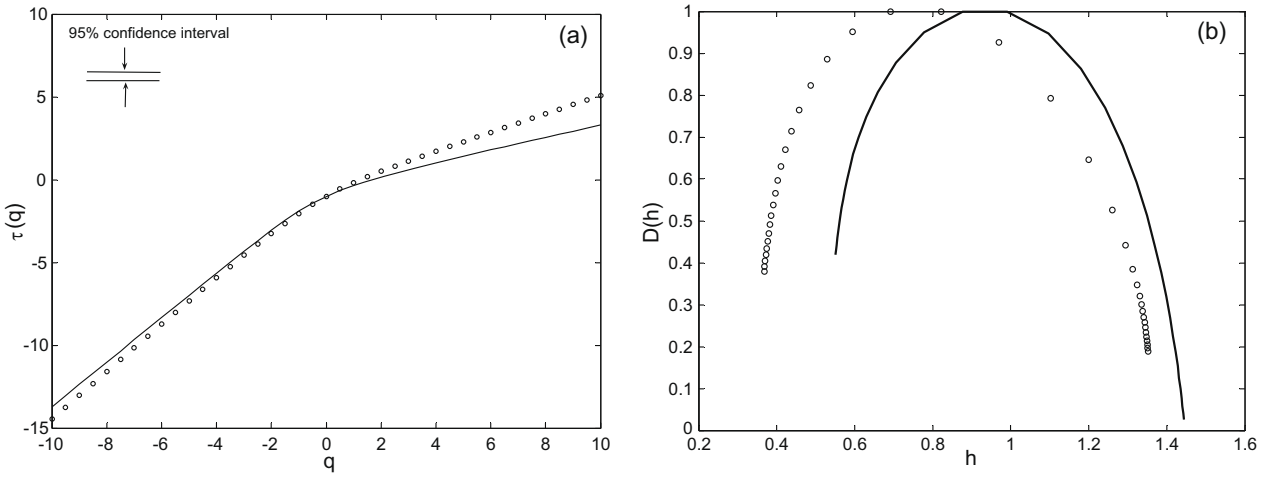

Fig. 15. Comparison of ROMS simulation results (solid curves) and satellite altimetry (white dots). (a) Scaling exponents. (b) Multi-fractal spectrum. Ninety-five percent confidence interval computed for the satellite altimetry data is shown in the upper left hand corner of the left panel.

Spectra of SSH anomaly fields reproduced by ROMS and those extracted from satellite altimetry (AVISO products) were similar on the most dominant spatial scales for a temporal scale ranging from 1 to 12 months. However, the spectrum extracted from the data decayed more quickly with k than that calculated from the ROMS simulation. This was explained by the fact that the AVISO product was smoothed at small scales to reduce measurement error near the coastline.

Multi-fractal features of the spectral entropies (scaling exponents and width of the multi-fractal spectrum) for ROMS and satellite altimetry were compared. By analysis using the scaling exponent, it was determined that both the entropies seemed to be very close for $-8 < q < 4$. This confirmed that large and small fluctuations in entropy behavior were scaled in a similar manner for both entropies. For both the entropies, $D(h)$ reached maximum value as $h > 0.5$, indicating the existence of long correlations in the entropy behavior.

To summarize, for the energy-dominant scales, ROMS simulation results generally demonstrate good agreement with observed RAFOS float and satellite observations for temporal scales of 1 to 12 months. Differences between ROMS and data, which were observed at some spatio-temporal scales, were due to the idealized (climatological) external forcing and open boundary conditions.

Finally, it is noted, that although ROMS has reproduced the observations within observational error, it is unlikely that the results of the ROMS simulations do not exist in nature. All estimates for the kinematic characteristics of the westward transport and statistical features of SSH anomaly field (time-averaged spectra, multi-fractal spectra and others) were robust to observational errors (at least for small-perturbations inserted in the data and the shortness of the length of observation series). The same property is the main characteristic of a robust dynamical regime (attractor) of a flow. Any model solution obtained with perturbations in external forcing or/and initial conditions should tend to the attractor with time [this property is called attractor stability (Kaneko, 1998)] and subsequently does not leave the attractor. It is likely that robust characteristics of the flow indicate the existence of an attractor. In this case, for any solution reproduced by ROMS within observational error, an attractor exists and therefore this solution is useful for scientific analysis of mesoscale/submesoscale variability off Central California.

Acknowledgements

This study was supported by NSF Award OCE-0530748 and OCE-0827168 for Prof. C. Collins and OCE-0827169 for Dr. L. Ivanov. Dr. T. Margolina thanks the National Research Council for a Research Associateship at the Naval Postgraduate School. Comments of Prof. J. McWilliams (University of California in Los Angeles, CA) were extremely useful and improved and clarified the manuscript. Thanks are also given to two anonymous reviewers for comments on the manuscript. MATLAB®'s Wavelet Toolbox and Fraclab (INRIA – http://fractales.inria.fr/) were used for multi-fractal analysis of spectral entropy series.

Appendix A

A.1. Boundary conditions for basis functions

The boundary conditions to produce modal decomposition (2) are briefly described. Necessary details of the approach can be found in Ivanov and Collins (in press).

A.1.1. Rigid boundary

Irrotational and rotational components of velocity are chosen such that the normal component of velocity along a rigid boundary was equal to zero. Therefore, SSH should be constant on the same boundary. This obviously is an approximate condition. However, error accompanying this approximate condition does not seem to affect ROMS-data comparisons. It quickly (exponentially) decays on small distances from the coastline and is small for distances larger than 30–50 km.

A.1.2. Open boundary

Here, the irrotational component of velocity should be equal to zero. This is an exact boundary condition. For the rotational component of velocity, the following approach was applied. The open boundary was moved along the external positive normal and a new computational domain of larger size (extended domain) was formed. Then, an approximate homogeneous boundary condition was introduced at the open boundary of the extended domain. The tangential component of velocity was assumed to be zero at the open boundary. This immediately results in zero normal derivatives for SSH at any point along the same boundary.

The basis functions were calculated in the extended domain and applied to approximate SSH within the original (non-extended) computational domain only. Errors of this approximation within this domain did not exceed 1%, and therefore were negligible.

Appendix B

B.1. Multi-fractal analysis of spectral entropy

The main disadvantage of using the structure participation function (4) to characterize the singular structure of entropy is that it often diverges for $q < 0$. A technique called the wavelet transform modulus maxima (WTMM) method (Muzzy et al., 1991) was used in order to avoid this effect. At a given scale, the partition function was calculated as a sum over local maxima of the modulus of the wavelet transform of the spectral entropy (W). The wavelet partition function is defined as

$$Z_q(a) = \sum_{l \in L(a)} \left[\sup_{a' > a} |W(a', x_l(a'))| \right]^q, \tag{B.1}$$

where $L(a)$ is a set of all the maxima lines l existing at a scale a, and $x_l(a)$ is the position, at a, of the maximum belonging to the line l. Each line $l = \{x_l(a), a\}$ is oriented (when a goes to 0) toward a point $b_l(0)$, which corresponds to a singularity of S.

Because one does not sum over places where the wavelet modulus is zero, the partition function is also defined for $q < 0$. If $S(t)$ is self-similar, then along the maxima line the partition function behaves as

$$Z_q(a) \sim a^{\tau(q)}, \tag{B.2}$$

A value of $\tau(q)$ is calculated for some fixed q and determines the slope of $\ln Z_q(a)(\ln a)$. By varying q in (B.1), the spectrum of the scaling exponents can be constructed. The Hölder exponent and multifractal spectrum $D(h)$ are,

$$h = d\tau/dq, \quad D(h) = qh - \tau(q) \tag{B.3}$$

respectively. If the Hölder exponent is constant, then $h = 0.5$, 1 and 1.5 corresponds to white noise, $1/f$ noise and a Weiner random process, respectively (Feder, 1988).

References

Artale, V., Boffetta, G., Celani, A., Cencini, M., Vulpiani, A., 1997. Dispersion of passive tracers in closed basins: beyond the diffusion coefficient. Phys. Fluids 9 (11), 3162–3171.

Aubry, N., Guyonnet, R., Lima, R., 1991. Spatio-temporal analysis of complex signals: theory and applications. J. Stat. Phys. 64 (3/4), 683–739.

Brink, K.H., Beardsley, R.C., Paduan, J., Limeburner, R., Caruso, M., Sires, J.G., 2000. A view of the 1993–1994 California Current based on surface drifters, floats, and remotely sensed data. J. Geophys. Res. 105 (C4), 8575–8604.

Capet, X., McWilliams, J.M., Molemaker, M.J., Shchepetkin, A.F., 2008. Mesoscale to submesoscale transition in the California Current system: frontal processes. J. Phys. Oceanogr. 38 (1), 44–64.

Cardoso, O., Gluckmann, B., Parcollet, O., Tabeling, E., 1996. Dispersion in a quasi-two-dimensional-turbulent flow: an experimental study. Phys. Fluids 8 (1), 209–214.

Centurioni, L.R., Ohlmann, J.M., Niiler, P.P., 2008. Permanent meanders in the California Current System. J. Phys. Oceanogr. 38, 1690–1710.

Chu, P.C., Ivanov, L.M., Kantha, L.H., Margolina, T.M., Melnichenko, O.V., Poberezhny, Y.A., 2004a. Lagrangian predictability of high-resolution regional ocean models: the special case of the Gulf of Mexico. Nonlinear Proc. Geophys. 10, 1–20.

Chu, P.C., Ivanov, L.M., Margolina, T.M., 2004b. Rotation method for reconstructing process and field from imperfect data. J. Bifur. Chaos 14 (8), 2991–2997.

Collecte Localisation Satellites, 2006. SSALTO/DUACS user handbook: (M)SLA and (M)ADT near-real time and delayed time products, CLS-DOS-NT-06.034, Ramonville Saint-Agne, France, 54pp.

Collins, C.A., Ivanov, L.M., Melnichenko, O.V., Garfield, N., 2004. California Undercurrent variability and eddy transport estimated from RAFOS float observations. J. Geophys. Res. 109, C05028. doi: 10.1029/2003JC002192.

Corrsin, S., 1974. Limitations of gradient transport models in random walks and in turbulence. Adv. Geophys. 18a, 25–60.

d'Ovidio, F., Fernandez, V., Hernandez-Garsia, E., Lopez, C., 2004. Mixing structures in the Mediterranean Sea from finite-size Lyapunov exponents. Geophys. Res. Lett. 31, L17203. doi: 10.1029/2004GL.020328.

Da Silva, A., Young, C., Levitus, S., 1994. Atlas of Surface Marine Data 1994, vols. 1–5. NOAA Atlas NESDIS 6-10, U.S. Government Printing Office.

Daubechies, I., 1992. Ten Lectures on Wavelets, Soc. Ind. Appl. Math., Philadelphia, 357pp.

Davis, R.E., 1991. Observing the general circulation with floats. Deep Sea Res. 38 (Suppl. 1), S531–S571.

Ducet, N., Le Traon, P.Y., Reverdin, G., 2000. Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2. J. Geophys. Res. 105, 19477–19498.

Feder, J., 1988. Fractals. Plenum Press, NY. 283pp.

Furrer, R., Sain, S.S., Nychka, D., Meehl, G.A., 2007. Multivariate Bayesian analysis of atmosphere-ocean general circulation models. Environ. Ecol. Stat. 14 (3), 249–266.

Gardiner, C.W., 2004. Handbook for Stochastic Methods for Physics, Chemistry and the Natural Sciences. Springer. 442pp.

Garfield, N., Collins, C.A., Paquette, R.G., Carter, E., 1999. Lagrangian exploration of the California Undercurrent, 1992–1995. J. Phys. Oceanogr. 29, 560–583.

Garraffo, Z., Mariano, A.J., Griffa, A., Veneziani, C., Chassignet, E., 2001. Lagrangian data in a high resolution numerical simulation of the North Atlantic. I: Comparison with *in-situ* drifter data. J. Mar. Syst. 29, 157–176.

Good, P.I., 1996. Re-Sampling Methods. A Practical Guide to Data Analysis. Birkhauser, Boston, Basel, Berlin. 238pp.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., Wallis, J.R., 1979. Probability-weighted moments: definition and relation to parameters of several distributions expressible in inverse form. Water Resour. Res. 15, 1049–1054.

Hosking, J.R.M., Wallis, J.K., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press, Cambridge. 224pp.

Ivanov, L.M., Chu, P.C., 2007a. On stochastic stability of regional ocean models with uncertainty in wind forcing. Nonlinear Proc. Geophys. 14 (5), 655–670.

Ivanov, L.M., Chu, P.C., 2007b. On stochastic stability of regional ocean models to finite-amplitude perturbations of initial conditions. Dyn. Atmos. Oceans 43 (3–4), 199–225.

Ivanov, L.M., Collins, C.A., in press. Modal decomposition of oceanic circulation: applications for high-resolution models and Lagrangian data. In: New Ideas in Oceanic Circulation, Nova Science Publishers, Inc., New York.

Ivanov, L.M., Kirwan, A.D., Melnichenko, O.V., 1994. Prediction of the stochastic behavior of nonlinear systems by deterministic models as a classical time-passage probabilistic problem. Nonlinear Proc. Geophys. 1, 224–233.

Ivanov, L.M., Collins, C.A., Margolina, T.M., Piterbarg, L.I., Eremeev, V.N., 2008. On westward transport processes off Central California revealed by RAFOS floats. Geophys. Res. Lett. 35, L18604. doi: 10.10291/2008GL034689.

Ivanov, L.M., Collins, C.A., Margolina, T.M., 2009. System of quasi-zonal jets off California extracted from satellite altimetry, Geophys. Res. Lett. 36, L03609, doi:10.1029/2008GL036327.

Joliffe, I.T., Stephenson, D.B., 2003. Forecast Verification: A Practitioner's Guide in Atmospheric Science. John Wiley & Son. 240pp.

Kaneko, K., 1998. On the strength of attractors in a high-dimensional system. Milnor attractor: Milnor attractor network, robust global attraction, and noise-induced selection. Physica D 124, 322–344.

Kumar, P., Foufoula-Georgiou, E., 1997. Wavelet analysis for geophysical applications. Rev. Geophys. 34 (4), 385–412.

La Casce, J.H., Bower, A., 2000. Relative dispersion in the subsurface North Atlantic. J. Mar. Res. 58, 863–894.

Lacorata, G., Aurell, E., Vulpiani, A., 2001. Drifter dispersion in the Adriatic Sea: Lagrangian data and chaotic model. Ann. Geophys. 19, 1291–1299.

Griffa, A., Kirwan, A.D., Mariano, A.J., Ozgokmen, T.M., Rossby, T. (Eds.), 2007. Lagrangian Analysis and Prediction of Coastal and Ocean Dynamics (LAPCOD). Cambridge University Press, Cambridge.

Le Traon, P.Y., Dibarboure, G., 1999. Mesoscale mapping capabilities of multiple-satellite altimeter missions. J. Atmos. Ocean. Technol. 16, 1208–1223.

Le Traon, P., Faugere, Y., Hernandez, F., Dorandeu, J., Mertz, F., Albain, M., 2003. Can we merge GEOSAT Follow-On with TOPEX/POSEIDON and ERA-2 for an improved description of the ocean circulation? J. Atmos. Ocean. Technol. 20 (6), 889–895.

Mackas, D.L., 2005. Interdisciplinary oceanography of the western North American continental margin. In: Brink, K.H., Robinson, A.R., Chapter 14. The Sea, vol. 12. Harvard University Press.

Mallatt, S., 1989. A theory for multi-resolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11, 674–693.

Mallatt, S., Zhong, S., 1992. Characteristics of signals from multiscale edges. IEEE Trans. Pattern Anal. Mach. Intell. 14, 710–732.

Marchesiello, P., McWilliams, J.C., Shchepetkin, A., 2003. Equilibrium structure and dynamics of the California Current System. J. Phys. Oceanogr. 33, 753–783.

Margolina, T.M., Collins, C.A., Rago, T.A., Paquette, R.G., Garfield, N., 2006. Intermediate level Lagrangian subsurface measurements in the Northeast Pacific: isobaric RAFOS float data. Geochem. Geophys. Geosys. 7, 009002. doi: 10.10291.GC00.1295.

McClean, J.L., Poulain, P.-M., Pelton, J.W., 2002. Eulerian and Lagrangian statistics from surface drifters and a high-resolution POP simulation in the North Atlantic. J. Phys. Oceanogr. 32, 2473–2491.

Miller, A.J., McWilliams, J.C., Schneider, N., Allen, J.S., Barth, J.A., Beardsley, R.C., Chavez, F.P., Chereskin, T.K., Edwards, C.A., Haney, R.L., Kelly, K.A., Kindle, J.C., Ly, L.N., Moisan, J.R., Noble, M.A., Niiler, P.P., Oey, L.Y., Schwing, F.B., Shearman, R.K., Swenson, M.S., 1999. Observing and modeling the California Current System, EOS. Trans. Am. Geophys. Union 80, 533–539.

Monin, A.S., Yaglom, A.M., 1971. Statistical Fluid Mechanics: The Mechanics of Turbulence. MIT Press, Cambridge, Mass. 784pp.

Muzzy, J.F., Bacry, E., Arnedo, A., 1991. Wavelets and multi-fractal formalism for singular signals: application to turbulent data. Phys. Rev. Lett. 67, 3515–3518.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation and confirmation of numerical models in the Earth Sciences. Science 263 (5147), 641–646.

Pasquero, C., 2005. Differential eddy diffusion of biogeochemical tracers. Geophys. Res. Lett. 32, L17603. doi: 10.1029/2005GL023662.

Pavlov, A.N., Anishenko, V.S., 2007. Multi-fractal analysis of complex signals. Usp. Fiz. Nauk 177 (8), 859–876.

Rossby, T., Dorson, D., Fontaine, J., 1986. The RAFOS system. J. Atmos. Ocean. Technol. 3, 672–679.

Shchepetkin, A.F., McWilliams, J.C., 2003. A method for computing horizontal pressure-gradient force in an ocean model with a non-aligned vertical coordinate. J. Geophys. Res. 108 (C3), 35.1–35.4.

Shchepetkin, A.F., McWilliams, J.C., 2005. Regional Ocean Model System: a split-explicit ocean model with a free-surface and topography-following vertical coordinate. Ocean Modell. 9, 347–404.

Shchepetkin, A.F., McWilliams, J.C., 2008. Computational kernel algorithms for fine-scale, multi-process, long-term oceanic simulations. In: Teman, R., Tribbia, J. (Eds.), Handbook of Numerical analysis on Computational Methods for the Oceans and Atmosphere, pp. 56–79.

Sornette, D., Davis, A.B., Ide, K., Vixie, K.R., Pisarenko, V., Kamm, J.K., 2007. Algorithm for model validation: theory and applications. Proc. Natl. Acad. Sci. USA 104 (16), 6562–6567.

Thomson, D.J., 1984. Random walk modeling of diffusion in inhomogeneous turbulence. Q. J. Roy. Meteor. Soc. 110 (466), 1107–1120.

Wunsch, C., Heimbach, P., 2007. Practical global oceanic state estimation. Physica D 230, 197–208.